

AD-A166 139

WALSH-FOURIER ANALYSIS OF DISCRETE-VALUED TIME SERIES
(U) PITTSBURGH UNIV PA CENTER FOR MULTIVARIATE ANALYSIS
D S STOFFER NOV 85 TR-85-46 AFOSR-TR-86-0064

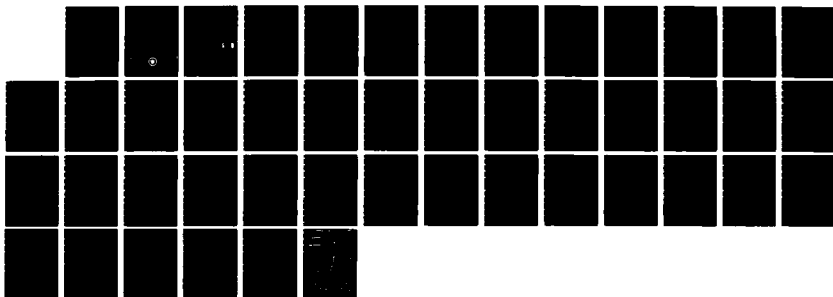
1/1

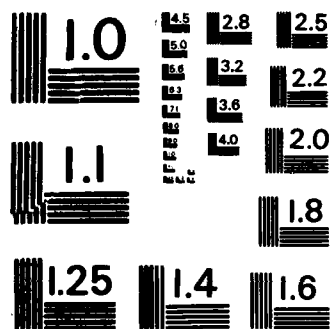
UNCLASSIFIED

F49620-85-C-0008

F/G 12/1

NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

WALSH-FOURIER ANALYSIS OF DISCRETE-
VALUED TIME SERIES

by

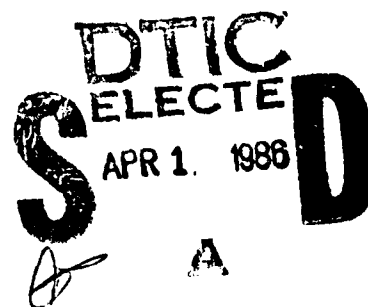
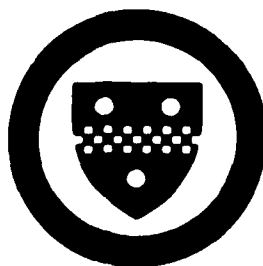
David S. Stoffer *

University of Pittsburgh

AD-A166 139

Center for Multivariate Analysis

University of Pittsburgh



Approved for public release;
distribution unlimited.

DTIC FILE COPY

86 4 1 012

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)
NOTICE OF TRANSMITTAL TO DTIC
This technical report has been reviewed and is
approved for public release in accordance with AFR 190-12.
Distribution is unlimited.
MATTHEW J. KEMPER
Chief, Technical Information Division

WALSH-FOURIER ANALYSIS OF DISCRETE-
VALUED TIME SERIES

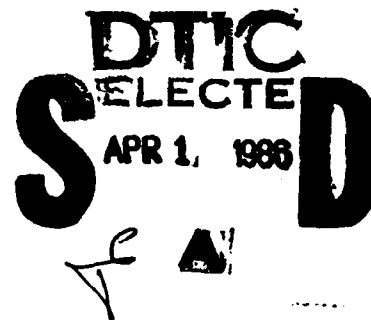
by

David S. Stoffer *

University of Pittsburgh

November 1985

Technical Report 85-46



Center for Multivariate Analysis
Fifth Floor, Thackeray Hall
University of Pittsburgh
Pittsburgh, PA 15260

*Part of the work of this author was supported by the Air Force Office of Scientific Research under Contracts F49620-85-C-0008 and AFOSR-84-0113. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

(A)

ABSTRACT

An approach to the analyses of discrete-valued time series is discussed. The analyses are accomplished in the spectral domain using the Walsh-Fourier transform which is based on Walsh functions. This approach will enable an investigator of discrete systems to analyze the data in terms of square-waveforms and sequency rather than sine-waves and frequency.

This document
~~We develop~~ a general signal-plus-noise type model for discrete-valued time series in which Walsh-Fourier spectral analysis is of interest. *The error* ~~We con-~~ sider the problems of detecting whether or not a common signal exists in repeated measures on discrete-valued time series and in discrete-valued processes collected in an experimental design. *It is* ~~We show~~ that these models may depend on unknown regression parameters and we develop consistent estimates of these parameters based on the finite Walsh-Fourier transform. *are developed* Applications to certain Markov models are given, however, the methods presented also apply to non-Markov cases.

Request for	
File	<input checked="checked" type="checkbox"/>
Dist	<input type="checkbox"/>
Un. processed	<input type="checkbox"/>
Classification	
Distinction/	
Availability Codes	
Dist	Avail and/or Special
<i>A-1</i>	

1. INTRODUCTION

Implicit in the Fourier (trigonometric) analysis of time series is one of two extreme assumptions about the process: (1) The very long stretch of the time series is the only time series we want to consider and consists of the superposition of not too many sinusoidal terms of substantially different frequencies; and (2) The time series is to be regarded as a realization of an ergodic Gaussian process. It is one of many possible time series and the analyses are directed toward the properties of the ensemble of the series, not toward those of a specific realization (cf. Brillinger and Tukey (1982)).

There are, however, many physical situations in which time series are either positive or discrete and are patently non-normal so that the analyses cannot be handled by transforming the data and applying Gaussian techniques (see, for example, Lewis (1980), p. 154). Similarly there are processes, such as those which take values in a discrete finite set, which can neither be thought of as Gaussian, nor as the superpositions of well separated sinusoids. Models for discrete-valued time series which have an ARMA structure are considered in Jacobs and Lewis (1978 a,b, 1983) and in Lewis (1980). For the case of continuous-valued non-normal time series it is perhaps still reasonable, in appropriate cases, to do spectral analysis via trigonometric methods. However, in the cases where time series take values in a discrete (and possibly finite) set, it makes little statistical sense to correlate the data with sines and cosines. As an alternative, we suggest that the spectral analysis of discrete-valued time series be accomplished in the "sequency" domain via the Walsh-Fourier transform (cf. Ahmed and Rao (1975), Kohn (1980 a) or Morettin (1981)). This seems to be a natural alternative to the usual Fourier analysis since the Walsh-Fourier transform is based on the "square-wave" Walsh functions. This approach would enable investigators to analyze discrete-valued time series (which we may think of as square-waveforms) in terms

of square-waves and sequency (switches per unit time) rather than sine-waves and frequency. As empirically demonstrated in Beauchamp (1975), "the respective roles of Walsh and Fourier spectral analysis for discontinuous and smooth-varying waveforms are clear. Where the signal is derived from sinusoidally-based waveforms, Fourier analysis is relevant. Where the signal contains sharp discontinuities and a limited number of levels, Walsh analysis is appropriate".

The Walsh functions, which are defined via the Rademacher functions (cf. Ahmed and Rao (1975), Kohn (1980 a) or Morettin (1981)), form a complete orthonormal sequence on $[0,1)$ and take on only two values, $+1$ and -1 (or 'on' and 'off'). They are ordered by the number of zero-crossings (or switches) which is called sequency. Let $W(n,\lambda)$, $n = 0,1,2,\dots$, $0 \leq \lambda < 1$, denote the n -th sequency-ordered Walsh function, then $W(n,\cdot)$ makes n zero-crossings in $[0,1)$. The first eight discrete, sequency-ordered Walsh functions, $W(n,m/N)$, $m,n = 0,1,\dots,7$, corresponding to a sample of length $N = 2^3$ are shown in Figure 1.1 in an 8×8 symmetric matrix called the Walsh ordered Hadamard matrix, $H_W(3)$. See Appendix A for details. We note that other orderings exist, for example, Paley order and Hadamard order are often used (cf. Ahmed and Rao (1975)), however, sequency or Walsh ordering is more natural in that it is comparable to the frequency ordering of sines and cosines. We will discuss methods of generating the discrete Walsh functions in Appendix A.

$$H_W(3) = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \end{bmatrix}$$

FIGURE 1.1: Sequence-ordered discrete Walsh functions, $W(n,m/N)$, $n,m = 0,1,\dots,7$, for a sample of size $N = 2^3$ as the rows of a Hadamard matrix.

Walsh spectral analysis has been used for several purposes, primarily in the Engineering sciences, such as speech processing, word recognition, image coding and transmission, filtering and multiplexing. It has also been used to describe biological and medical systems such as monitoring EEG and ECG signals. See, for example, the Proceedings on the Applications of Walsh Functions, Ahmed and Rao (1975), Beauchamp (1975) and Harmuth (1972) to mention a few. Applications of Walsh functions in the statistics literature are rather scarce and we mention two. Ott and Kronmal (1976) use the Walsh transform in classification and prediction problems for strictly stationary binary time series. Stoffer and Panchalingam (1985) analyze simulated and real binary time series in the sequence domain.

At present, there are two modes of development of Walsh spectral analysis in the literature. The first mode is termed Walsh spectral analysis and is developed via the concept of dyadic stationarity. That is, it is based on processes $\{X(n); n = 0,1,2,\dots\}$ for which $\text{Cov}(X(n), X(n \oplus m)) = B(m)$ is a

function only of the dyadic distance between n and $n \oplus m$, where $n \oplus m$ denotes the dyadic addition of n and m (cf. Morettin (1974 b, 1981) for definitions, discussions and references). In this mode, one has in mind that the process of interest is the superposition of not too many Walsh functions of substantially different sequences, that is, $X(n) = \sum_{k=1}^K Z(k)W(n, \lambda_k)$ where $Z(1), \dots, Z(K)$ are uncorrelated random variables with mean zero and variance σ_k^2 , $k = 1, \dots, K$, with $\lambda_1, \dots, \lambda_K$ constants, $\lambda_i \neq \lambda_j$, $i \neq j$. The other mode of development is termed Walsh-Fourier spectral analysis and is based on real time stationarity. Theoretical results concerning the statistical application of Walsh-Fourier spectral analysis are relatively recent and to the best of my knowledge, are limited to the works of Kohn (1980 a,b), Morettin (1974 a, 1981, 1983) and Stoffer (1985). One is warned that in the examination of non-mathematical literature concerning this subject, to keep in mind that the two different modes exist, although the particular mode is not always apparent. This matter can be quite confusing since the results are considerably different and results from one mode do not typically apply to the other. An excellent review of the two different approaches is given in Morettin (1981). We believe that although dyadic time has some theoretical appeal in the Walsh spectral domain, due to its strange behavior (see, for example, the discussions in Robinson (1972) and Beauchamp (1975)), it is of little practical use. We therefore, concentrate on real time stationary processes.

A brief account of the existing Walsh-Fourier theory as well as some new results and necessary tools to be used in the sequel are given in the next section. In Section 3 we present a general signal-plus-noise type model for discrete-valued time series in which Walsh-Fourier spectral analysis is of interest. Next we consider the problem of detecting whether or not a common signal exists in repeated measurements on discrete-valued time series. It is

shown that this method extends to discrete-valued time series collected in an experimental design. Then, we consider the analysis of discrete signal-plus-noise models in a regression setting in which the signal is observable but the process depends on regression parameters. Our main goal is to estimate, via Walsh-Fourier spectral methods, the regression parameters and the noise spectrum. In all cases we give concrete examples by showing that certain Markov chains satisfy our models, however, the methods presented here also apply in non-Markov cases. For completeness, we provide a discussion in Appendix A on simple and economic methods for generating the discrete sequency-ordered Walsh functions and hence the finite Walsh-Fourier transform.

2. PRELIMINARIES

In this section we give definitions, establish some new results, and give a brief summary of the existing Walsh-Fourier theory for stationary time series which we use in the sequel. In particular, we concentrate on those properties which we can directly apply to discrete-valued time series. Hence, for example, we do not consider properties of processes which are generalized linear processes; such processes can be handled by the theory given in Kohn (1980 a,b), Morettin (1983) and Stoffer (1985). At present, our discussion will be for univariate time series, the multivariate versions follow in an obvious way and we mention them briefly at the end of this section.

Let $X(0), X(1), \dots, X(N-1)$ be a sample of length $N = 2^p$, $p > 0$ integer, from a weakly stationary time series, $\{X(n), n = 0, \pm 1, \pm 2, \dots\}$, with absolutely summable autocovariance function, $\gamma(h) = \text{Cov}(X(n), X(n+h))$, $h = 0, \pm 1, \pm 2, \dots$. We assume for now that the constant mean value of $X(n)$ is zero. Let $W(n, \lambda)$ be the n -th Walsh function in sequence order, and let

$$d_N(\lambda) = N^{-1/2} \sum_{n=0}^{N-1} X(n)W(n, \lambda), \quad 0 \leq \lambda < 1 \quad (2.1)$$

be the finite (or discrete) Walsh-Fourier transform of the data. The logical covariance of $X(n)$ (cf. Robinson (1972) or Kohn (1980 a)) is defined to be

$$\tau(j) = N^{-1} \sum_{k=0}^{N-1} \gamma(j \oplus k - k)$$

where by $j \oplus k$ we mean the dyadic addition of j and k . It can then be shown (cf. Kohn (1980 a)) that the variance of $d_N(\lambda)$ is given by

$$\text{Var}\{d_N(\lambda)\} = \sum_{j=0}^{N-1} \tau(j)W(j,\lambda). \quad (2.2)$$

Taking the limit $(N \rightarrow \infty)$ in (2.2) we have that $\text{Var}\{d_N(\lambda)\} \rightarrow f(\lambda)$ where

$$f(\lambda) = \sum_{j=0}^{\infty} \tau(j)W(j,\lambda), \quad 0 \leq \lambda < 1 \quad (2.3)$$

is called the Walsh-Fourier spectral density of $X(n)$. We note that $f(\lambda)$ exists since the absolute summability of $\gamma(h)$ implies the absolute summability of $\tau(j)$. Specifically, Kohn (1980 a, Lemma 3) shows that if

$$\lim_{n \rightarrow \infty} \sum_{|j| < 2^n} (1 - |j|/2^n) |\gamma(j)| < \infty \quad (2.4)$$

then $\sum_{j=0}^{\infty} |\tau(j)| < \infty$ and $f(\lambda)$ is well-defined.

If $X(0), X(1), \dots, X(N-1)$ is a sample of length $N = 2^p$, the finite transform (2.1) is calculated for $\lambda_N = m/N$, $m = 0, 1, \dots, N-1$. Since the discrete Walsh functions are symmetric in their arguments for $N = 2^p$, that is,

$$W(n, m/N) = W(m, n/N), \quad m, n = 0, 1, \dots, N-1, \quad (2.5)$$

the value of λ_N in the finite Walsh-Fourier transform corresponds to sequency. As with the usual Fourier analysis, if the mean of the series is unknown, the only sequency of the form $\lambda_N = m/N$ for which the transform cannot be evaluated is at the zero ($m = 0$) sequency. To see this, let $\theta = EX(n)$, all n , and note that for $m = 0, 1, \dots, N-1$,

$$N^{-1} \sum_{n=0}^{N-1} W(n, m/N) = \delta_0^m. \quad (2.6)$$

where δ is the Kronecker delta. See Kohn (1980 a, Lemma 1). It is clear from (2.6) that the mean centered transform will be the uncentered transform except at $m = 0$, and in particular

$$E\{d_N(m/N)\} = N^{-1/2} \sum_{n=0}^{N-1} \theta W(n, m/N) = N^{1/2} \theta \delta_0^m,$$

$m = 0, 1, \dots, N-1$.

Kohn (1980 a, Corollary 3) gives the following useful results on the convergence of the second moment of the finite Walsh-Fourier transform under condition (2.4). Let λ_N be dyadically rational (that is, its binary representation is finite). If $\lambda_N \oplus \lambda \rightarrow 0$ as $N = 2^p \rightarrow \infty$, then

$$E\{d_N^2(\lambda_N)\} \rightarrow f(\lambda). \quad (2.7)$$

In general, the asymptotic covariance of the Walsh-Fourier transform at two distinct sequences is not zero (cf. Kohn (1980 a), Theorem 3). However, if $\lambda_{1,N}$ and $\lambda_{2,N}$ are dyadically rational and $|\lambda_{1,N} - \lambda_{2,N}| \geq N^{-1}$ with $\lambda_{i,N} \oplus \lambda \rightarrow 0$, $i = 1, 2$ as $N = 2^p \rightarrow \infty$, then

$$E\{d_N(\lambda_{1,N})d_N(\lambda_{2,N})\} \rightarrow 0.$$

Various authors have established central limit theorems for the finite Walsh-Fourier transform under a wide range of conditions (cf. Kohn (1980 a), Morettin (1983) and Stoffer (1985)). We state three versions which are applicable to discrete-valued time series. The first version (Assumption 2.1) follows its trigonometric counterpart given in Hannan (1973) and can be found in Kohn (1980 a, Theorem 4). The second version (Assumption 2.2) follows its trigonometric counterpart based on the existence of higher moments given in Brillinger

(1975) and can be found in Morettin (1983, Theorem 1). We remark that the above two versions exist side by side and that neither is included in the other (cf. Morettin (1983)). The third version (Assumption 2.3) is similar to Stoffer (1985, Theorem 3.1), however it is sufficiently different to warrant a proof which is provided in Appendix B. This version establishes a central limit theorem for discrete-valued second-order stationary processes which satisfy a type of finite dependence property.

ASSUMPTION 2.1: $X(n)$ is strictly stationary with zero mean. Let F_n be the σ -field generated by $\{X(j), j \leq n\}$; put

$$\alpha_j = [E\{E(X(n) | F_{n-j}) - E(X(n) | F_{n-j-1})\}^2]^{1/2}, \quad j \geq 0,$$

$F_{-\infty}$ is trivial and $\sum_{j=0}^{\infty} \alpha_j < \infty$.

ASSUMPTION 2.2: $X(n)$ is strictly stationary with zero mean and finite moments.

Let $C_r(j_1, \dots, j_r) = \text{Cum}\{X(j_1), \dots, X(j_r)\}$ be the r -th cumulant of $X(n)$,

$$j_1, \dots, j_r = 0, \pm 1, \pm 2, \dots; \quad \sum_{j_1=0}^{\infty} \dots \sum_{j_{r-1}=0}^{\infty} |C_r(j_1, \dots, j_{r-1})| < \infty.$$

ASSUMPTION 2.3: $X(n)$ is second-order stationary with zero mean and covariance function $\gamma(k)$. $\sup_n E\{|X(n)|^{2+\delta}\} < \infty$ for some $\delta > 0$. Let F_n be the σ -field generated by $\{X(j), j \leq n\}$. There exists a positive integer κ such that (i) $E\{X(n) | F_{n-\kappa}\} = 0$ a.s., and (ii) $E\{X(n)X(n+k) | F_{n-\kappa}\} = \gamma(k)$ a.s. for $k = 0, 1, \dots, \kappa-1$.

If condition (2.4) and either Assumption 2.1, 2.2 or 2.3 hold, then $d_N(\lambda)$

converges in distribution to a normal variate with mean zero and variance $f(\lambda)$ given by (2.3). We note, for example, that discrete-valued time series which are based on mixtures, such as the discrete MA processes described in Jacobs and Lewis (1978 a,b, 1983), Lewis (1980), and the geometric processes discussed in Langberg and Stoffer (1985), satisfy the dependence properties of Assumption 2.3.

In order to be able to consistently estimate the Walsh-Fourier spectrum, we need asymptotic results for smoothing the Walsh-Fourier periodogram, $I_N(\lambda_N) = d_N^2(\lambda_N)$. We now state the following results in a theorem for use in the sequel. The theorem combines the three previous central limit theorems with a theorem given in Kohn (1980 a, Theorem 4).

THEOREM 2.1: Let condition (2.4) and either Assumption 2.1, 2.2 or 2.3 hold.

Let $\lambda_{j,N} = j/N$, $1 \leq j \leq N-1$, and suppose for $\{\lambda_{j(1),N}, \dots, \lambda_{j(M),N}\}$, $\lambda_{j(m),N} \otimes \lambda \rightarrow 0$ as $N \rightarrow \infty$, $m = 1, \dots, M$ and $|\lambda_{j(\ell),N} - \lambda_{j(k),N}| \geq N^{-1}$ for $\ell \neq k$, $\ell, k = 1, \dots, M$. Then $\underset{\mathcal{D}}{d}_N \rightarrow N(0, \Delta)$ where $\underset{\mathcal{D}}{d}_N = (d_N(\lambda_{j(1),N}), \dots, d_N(\lambda_{j(M),N}))'$ and Δ is an $M \times M$ diagonal matrix with $f(\lambda)$ along the diagonal. Also

$$\underset{\mathcal{D}}{d}_N' \underset{\mathcal{D}}{d}_N \xrightarrow{\mathcal{D}} f(\lambda) \chi_M^2 \quad (2.8)$$

so that $M^{-1} \underset{\mathcal{D}}{d}_N' \underset{\mathcal{D}}{d}_N$ is an estimate of $f(\lambda)$ having variance $2f^2(\lambda)/M$.

If we let $M \rightarrow \infty$ as $N \rightarrow \infty$ with $M/N \rightarrow 0$ in Theorem 2.1, the smoothed periodogram $M^{-1} \underset{\mathcal{D}}{d}_N' \underset{\mathcal{D}}{d}_N$ is a mean square consistent estimate of the Walsh-Fourier spectrum $f(\lambda)$, $0 < \lambda < 1$.

Results obtained for the univariate finite Walsh-Fourier transform carry over, in an obvious way, to the vector case, say $\underline{X}(n) = (X_1(n), \dots, X_p(n))'$, except

that (cf. Kohn (1980 b), Section 3)

- (1) The logical covariance will now be $r \times r$ matrices given by

$$\tau(j) = N^{-1} \sum_{k=0}^{N-1} \{\Gamma(j \oplus k - k) + \Gamma'(j \oplus k - k)\} \quad (2.9)$$

where $\Gamma(h)$ is the $r \times r$ autocovariance matrix of $\underline{X}(n) = (X_1(n), \dots, X_r(n))'$.

- (2) The Walsh-Fourier spectrum $f(\lambda)$ is an $r \times r$ real positive semidefinite matrix.

$$(3) \quad \text{Cov}\{I_{ij}(\lambda) I_{lm}(\lambda)\} \rightarrow f_{il}(\lambda) f_{jm}(\lambda) + f_{im}(\lambda) f_{lj}(\lambda) \quad (2.10)$$

as $N \rightarrow \infty$ where $I_{ij}(\lambda)$ is the (i, j) -th element of the $r \times r$ periodogram matrix $I_N(\lambda) = g_N^X(\lambda) g_N^{X'}(\lambda)$ with

$$g_N^X(\lambda) = N^{-1/2} \sum_{n=0}^{N-1} \underline{X}(n) W(n, \lambda).$$

3. SIGNAL-PLUS-NOISE MODELS FOR DISCRETE-VALUED TIME SERIES

In this section we discuss models for discrete systems in which Walsh-Fourier analysis is desirable. We consider a discrete version of the signal-plus-noise models used for sinusoidal and for Gaussian processes. In general, write the $r \times 1$ vector, discrete-valued time series as

$$\underline{X}(n) = \underline{S}(n) + \underline{\varepsilon}(n), \quad n = 0, \pm 1, \pm 2, \dots \quad (3.1)$$

where $\underline{S}(n)$ is a random stationary discrete signal which possibly depends on unknown parameters $\underline{\theta} = (\theta_1, \dots, \theta_q)'$, and $\underline{\varepsilon}(n)$ is a zero-mean discrete-valued process (possibly white noise) which is uncorrelated with $\underline{S}(n)$. We note that the support of $\underline{X}(n)$, $\underline{S}(n)$ and $\underline{\varepsilon}(n)$ need not necessarily be the same and that there may be some dependence structure between $\underline{S}(n)$ and $\underline{\varepsilon}(n)$.

For a specific example of such a process, consider a macro model on a finite state space (cf. Basawa and Prakasa Rao (1980)). Let $X_j(n)$, $j = 1, \dots, r$ denote the number of individuals in state j at time n . In particular, $X_j(n)$ is the aggregate over several independent chains evolving simultaneously. Let θ_j , $j = 1, \dots, r$, be the probability of being in state j at any given time, and let Q denote the total number of individuals under consideration. Denote the $r \times 1$ vector $\underline{X}(n) = (X_1(n), \dots, X_r(n))'$ and suppose that $\underline{X}(n)$ is Markov with transition probabilities p_{ij} , $1 \leq i, j \leq r$. Then, for $n = 1, 2, \dots$,

$$EX_j(n) = \sum_{i=1}^r EX_i(n-1)p_{ij}, \quad 1 \leq j \leq r,$$

from which we obtain the signal-plus-noise model

$$X_j(n) = \sum_{i=1}^r X_i(n-1)p_{ij} + \varepsilon_j(n), \quad 1 \leq j \leq r, \quad (3.2)$$

where $\underline{\varepsilon}(n) = (\varepsilon_1(n), \dots, \varepsilon_r(n))'$, $n = 0, 1, 2, \dots$, is a zero-mean multinomial-type white noise process and is uncorrelated with the random signal

$\underline{S}(n) = (S_1(n), \dots, S_r(n))'$, $S_j(n) = \sum_{i=1}^r X_i(n-1)p_{ij}$, $1 \leq j \leq r$. To see that

$\underline{\varepsilon}(n)$ and $\underline{S}(n)$ are uncorrelated, first consider one individual at a time and

let $Y_{j,q}(n) = 1$ if individual q is in state j at time n , and $Y_{j,q}(n) = 0$

otherwise, $1 \leq q \leq Q$, $1 \leq j \leq r$. Denote the corresponding error term by

$w_{j,q}(n)$, that is, write $Y_{j,q} = \sum_{i=1}^r Y_{i,q}(n-1)p_{ij} + w_{j,q}(n)$. Then for $1 \leq q \leq Q$, $j, k = 1, \dots, r$,

$$\begin{aligned} \text{Cov}\{w_{j,q}(n), Y_{k,q}(n-1)\} &= \text{Cov}\{Y_{j,q}(n) - \sum_{i=1}^r Y_{i,q}(n-1)p_{ij}, Y_{k,q}(n-1)\} \\ &= \theta_k p_{kj} - \theta_k p_{kj} = 0. \end{aligned}$$

Noting that $\varepsilon_j(n)$ is the aggregate of the independent individual error terms,

$\varepsilon_j(n) = \sum_{q=1}^Q w_{j,q}(n)$, and that $S_j(n)$ is the aggregate of the independent

individual signal terms, $S_j(n) = \sum_{q=1}^Q \sum_{i=1}^r Y_{i,q}(n-1)p_{ij}$, we see that $\underline{\varepsilon}(n)$ and

$\underline{S}(n)$ are uncorrelated. Hence, we have exhibited a model of the discrete-valued signal-plus-noise form given in (3.1) where, in this case, the signal $\underline{S}(n)$ is a function of the r^2 transition probabilities, p_{ij} , $i, j = 1, \dots, r$.

For the signal-plus-noise model presented in this section, Walsh-Fourier analysis would be useful for detecting whether or not a discrete signal exists in a given system, and if so, determine the cyclic behavior, in terms of sequency, of the signal. Moreover, for discrete systems in which the signal is observable but the process depends on unknown parameters, Walsh-Fourier methods can be used to consistently estimate the parameters as well as to consistently estimate the

error spectrum. We discuss these types of analyses in the following subsections.

3.1 Detecting a Common Signal

Now, consider the problem of detecting whether or not a common discrete-valued signal exists in Q replications of a discrete-valued time series $\{X_q(n)\}$, $q = 1, \dots, Q$ which are of the signal-plus-noise form (3.1). Many of our techniques follow those of their trigonometric counterparts developed by Brillinger (1973; 1975; Section 7.9; 1980).

As an example, suppose we wish to analyze Q independent queueing situations hypothesized to be similar. Let $X_q(n)$, $q = 1, \dots, Q$, be the number of individuals in queue q at time n and suppose that $EX_q(n) = \theta_q$ for all n . Let $S(n)$ be a zero-mean stationary discrete-valued signal hypothesized to be common to all queues, such as a rate of change in the number of individuals in a queue, and let $\varepsilon_q(n)$, $q = 1, \dots, Q$ be stationary zero-mean discrete-valued processes with common Walsh-Fourier spectrum. Whether or not a common signal exists among the Q queueing situations is expressed in whether or not the signal, or equivalently its Walsh-Fourier spectrum is identically zero.

In general, we suppose that the discrete system is of the form $X_q(n)$, $q = 1, \dots, Q$; $n = 0, 1, \dots, N-1$, $N = 2^p$, $p > 0$ integer, and can be modeled as

$$X_q(n) = \theta_q + S(n) + \varepsilon_q(n) \quad (3.3)$$

where θ_q are constants, $S(n)$ is a realization of a stationary discrete-valued time series with mean zero, and $\varepsilon_q(n)$, $q = 1, \dots, Q$, are independent realizations of a zero-mean discrete-valued stationary time series which is uncorrelated with $S(n)$. Let $\gamma_{SS}(h)$ and $\gamma_{\varepsilon\varepsilon}(h)$, $h = 0, \pm 1, \pm 2, \dots$, denote the autocovariance functions of $S(n)$ and $\varepsilon_q(n)$, $1 \leq q \leq Q$, respectively. We

assume that $\gamma_{SS}(h)$ and $\gamma_{\epsilon\epsilon}(h)$ satisfy condition (2.4) and that every linear combination of $S(n)$ and $\epsilon_q(n)$, $1 \leq q \leq Q$, satisfies either Assumption 2.1, 2.2 or 2.3. Denote the logical covariances of $S(n)$ and $\epsilon_q(n)$, $1 \leq q \leq Q$ by $\tau_{SS}(j)$ and $\tau_{\epsilon\epsilon}(j)$, respectively, and the respective Walsh-Fourier spectrum of $S(n)$ and $\epsilon_q(n)$, $1 \leq q \leq Q$, by $f_{SS}(\lambda)$ and $f_{\epsilon\epsilon}(\lambda)$, $0 \leq \lambda < 1$. The following conditions hold:

- (1) $EX_q(n) = \theta_q$,
- (2) $\tau_{X_q X_q}(j) = \tau_{SS}(j) + \tau_{\epsilon\epsilon}(j)$,
- (3) $f_{X_q X_q}(\lambda) = f_{SS}(\lambda) + f_{\epsilon\epsilon}(\lambda)$,
- (4) $\tau_{X_q X_\ell}(j) = \tau_{SS}(j)$, $q \neq \ell$,
- (5) $f_{X_q X_\ell}(\lambda) = f_{SS}(\lambda)$, $q \neq \ell$.

Before we proceed with the analyses we need the following lemma.

LEMMA 3.1: Let $\{X_n\}$ and $\{Y_n\}$ be sequences of random variables on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ such that for all n , $EX_n^2 < \infty$ and $EY_n^2 < \infty$. If $X_n \xrightarrow{P} X$, $Y_n \xrightarrow{P} Y$, $EX_n^2 \rightarrow EX^2$, and $EY_n^2 \rightarrow EY^2$ as $n \rightarrow \infty$, then $EX_n Y_n \rightarrow EXY$.

PROOF: By Chung (1974, Theorem 4.5.4) the conditions $EX_n^2 < \infty$, $X_n \xrightarrow{P} X$, and $EX_n^2 \rightarrow EX^2$ imply the uniform integrability of $\{|X_n|^2\}$. Similarly, $\{|Y_n|^2\}$ is uniformly integrable. By the Cauchy-Swartz inequality,

$$\int_A |X_n Y_n| dP \leq \left[\int_A |X_n|^2 dP \right]^{1/2} \left[\int_A |Y_n|^2 dP \right]^{1/2}$$

for any set $A \in \mathcal{F}$ and hence by Chung (1974, Theorem 4.5.3), the uniform integrability of $\{|X_n|^2\}$ and $\{|Y_n|^2\}$ imply the uniform integrability of $\{|X_n Y_n|\}$. Clearly $X_n Y_n \xrightarrow{P} XY$ and $EX_n Y_n < \infty$, all n , hence by Chung (1974, Theorem 4.5.4), $EX_n Y_n \rightarrow EXY$.

The particular analyses of this section are based on the following theorem.

THEOREM 3.1: Let $\{X_q(n); q = 1, \dots, Q; n = 0, \dots, N-1\}$ satisfy the model conditions (3.3). Let $\lambda_{j(m),N}$, $m = 1, \dots, M$ be defined as in Theorem 2.1. Then the finite Walsh-Fourier transform $d_{N,q}(\lambda_{j(m),N}) = N^{-1/2} \sum_{n=0}^{N-1} X_q(n) W(n, \lambda_{j(m),N})$, $1 \leq q \leq Q$, has the representation

$$d_{N,q}(\lambda_{j(m),N}) = u_m + z_{q,m} \quad \text{a.s.} \quad (3.4)$$

as $N \rightarrow \infty$ where the u_m are independent $N(0, f_{SS}(\lambda))$ variates and the $z_{q,m}$ are independent $N(0, f_{\varepsilon\varepsilon}(\lambda))$ variates. Moreover, u_m and $z_{q,m}$ are mutually independent, $1 \leq m \leq M$, $1 \leq q \leq Q$.

PROOF: Taking the finite Walsh-Fourier transform of (3.3) we obtain

$$d_{N,q}(\lambda_{j(m),N}) = d_N^S(\lambda_{j(m),N}) + d_{N,q}^\varepsilon(\lambda_{j(m),N}), \quad 1 \leq q \leq Q$$

where $d_N^S(\cdot)$ and $d_{N,q}^\varepsilon(\cdot)$ are the Walsh-Fourier transforms of $S(n)$ and $\varepsilon_q(n)$, respectively. The constant term drops out in view of (2.6) and the fact that $\lambda_{j(m),N} \neq 0$. Invoking Skorokhods Representation Theorem (Skorokhod (1956)) and Theorem 2.1, $d_N^S(\lambda_{j(m),N})$ and $d_{N,q}^\varepsilon(\lambda_{j(m),N})$ have the representations

$d_N^S(\lambda_{j(m),N}) \rightarrow u_m$ a.s. and $d_{N,q}^E(\lambda_{j(m),N}) \rightarrow z_{q,m}$ a.s. as $N \rightarrow \infty$ where the u_m are independent $N(0, f_{SS}(\lambda))$ variates and the $z_{q,m}$ are independent $N(0, f_{\epsilon\epsilon}(\lambda))$ variates. It remains to show that u_m and $z_{q,m}$ are mutually independent and since they are jointly normal, it suffices to show that they are uncorrelated. Since $S(n)$ and $\epsilon_q(n)$ are uncorrelated we have that $E\{d_N^S(\lambda_{j(m),N})d_{N,q}^E(\lambda_{j(m),N})\} = 0$ for all N , $m = 1, \dots, M$, and $q = 1, \dots, Q$. Since $\gamma_{SS}(h)$ and $\gamma_{\epsilon\epsilon}(h)$ satisfy condition (2.4) we have by (2.7) $E\{d_N^S(\lambda_{j(m),N})^2\} \rightarrow f_{SS}(\lambda)$ and $E\{d_{N,q}^E(\lambda_{j(m),N})^2\} \rightarrow f_{\epsilon\epsilon}(\lambda)$ as $N \rightarrow \infty$, $1 \leq m \leq M$, $1 \leq q \leq Q$. Thus by Lemma 3.1 we have $E\{d_N^S(\lambda_{j(m),N})d_{N,q}^E(\lambda_{j(m),N})\} \rightarrow E\{u_m z_{q,m}\}$ as $N \rightarrow \infty$ and hence $E\{u_m z_{q,m}\} = 0$, $m = 1, \dots, M$, $q = 1, \dots, Q$.

Using the representation (3.4) we may proceed with the problem of determining whether or not a common discrete-valued signal exists. Our analysis follows that of the analysis of random effects models (cf. Scheffé (1959)). Consider the quantities

$$d_{N,\cdot}(\lambda_{j(m),N}) = Q^{-1} \sum_{q=1}^Q d_{N,q}(\lambda_{j(m),N}), \quad 1 \leq m \leq M$$

$$\sum_{m=1}^M d_{N,\cdot}^2(\lambda_{j(m),N}) \tag{3.5}$$

$$\sum_{q=1}^Q \sum_{m=1}^M [d_{N,q}(\lambda_{j(m),N}) - d_{N,\cdot}(\lambda_{j(m),N})]^2. \tag{3.6}$$

By Cochran's Theorem and in view of the representation (3.4), the quantities (3.5) and (3.6) have the almost sure representations

$$\sum_{m=1}^M (u_m + z_{\cdot, m})^2 = [f_{SS}(\lambda) + f_{\epsilon\epsilon}(\lambda)/Q] \chi_M^2 \quad \text{a.s.} \quad (3.7)$$

and

$$\sum_{q=1}^Q \sum_{m=1}^M (z_{q, m} - z_{\cdot, m})^2 = f_{\epsilon\epsilon}(\lambda) \chi_{M(Q-1)}^2 \quad \text{a.s.} \quad (3.8)$$

respectively, as $N \rightarrow \infty$, where the χ^2 variates in (3.7) and (3.8) are independent. The fact that the χ^2 variates are independent follows from the fact that $\text{Cov}(z_{q, m} - z_{\cdot, m}, z_{\cdot, m}) = 0$.

The test of the null hypothesis that there is no common signal, $S(n) \equiv 0$ or equivalently $f_{SS}(\lambda) \equiv 0$, may be examined for particular sequences λ by comparing

$$\frac{Q \sum_{m=1}^M d_{N, \cdot}^2(\lambda_{j(m), N})/M}{\sum_{q=1}^Q \sum_{m=1}^M [d_{N, q}(\lambda_{j(m), N}) - d_{N, \cdot}(\lambda_{j(m), N})]^2/M(Q-1)} \quad (3.9)$$

with an $F_{M, M(Q-1)}$ distribution. In view of (3.7) and (3.8), (3.9) has the representation

$$\frac{[f_{\epsilon\epsilon}(\lambda) + Q f_{SS}(\lambda)]}{f_{\epsilon\epsilon}(\lambda)} F_{M, M(Q-1)} \quad \text{a.s.}$$

as $N \rightarrow \infty$.

Using the representations (3.7) and (3.8) it is easy to find an asymptotically ($N \rightarrow \infty$) unbiased estimate of the Walsh-Fourier spectrum $f_{SS}(\lambda)$ of the unobserved signal, say $\hat{f}_{SS}(\lambda)$, $0 < \lambda < 1$. Denoting the numerator of (3.9) by $Q \hat{f}_{\bar{X}\bar{X}}(\lambda)$ and the denominator of (3.9) by $\hat{f}_{\epsilon\epsilon}(\lambda)$ we see that the desired

quantity is

$$\hat{f}_{SS}(\lambda) = \hat{f}_{\bar{X}\bar{X}}(\lambda) - Q^{-1} \hat{f}_{\epsilon\epsilon}(\lambda). \quad (3.10)$$

We note that due to the invariance properties of maximum likelihood estimates (MLEs) and the fact that the asymptotic distributions of the transforms are Gaussian, the estimate $\hat{f}_{SS}(\lambda)$ in (3.10) is also the bias corrected MLE of $f_{SS}(\lambda)$ is an asymptotic ($N \rightarrow \infty$) sense.

Next, we consider a one-way design. Suppose now that L different treatments are replicated on experimental units and stationary discrete-valued time series are recorded for each of the units. For example suppose L treatments are given to groups of patients (Q patients in each group) who suffer from a particular allergy, and we record the number of allergic reactions per day of each patient for N days. The data collected will not only be discrete count data, but for many allergies it is mostly on the order of 0,1,2,3 and very seldom above this. Hence, a Walsh-Fourier as opposed to trigonometric decomposition of the observations would be more useful. This type of design is similar to the balanced one-way random effects design considered in Scheffé (1959, Section 7.2) and in Brillinger (1980, Section 4.2). Specifically, for $q = 1, \dots, Q$, $\ell = 1, \dots, L$, $n = 0, 1, \dots, N-1$, $N = 2^p$, $p > 0$ interger, let

$$X_{q\ell}(n) = \theta_{q\ell} + S_{\ell}(n) + \epsilon_{q\ell}(n) \quad (3.11)$$

where $\theta_{q\ell}$ are constants; $S_{\ell}(n)$ are L independent stationary discrete-valued time series with mean zero, logical covariance $\tau_{SS}(j)$ and Walsh-Fourier spectrum $f_{SS}(\lambda)$; and $\epsilon_{q\ell}(n)$ are independent realizations of a zero-mean discrete-valued stationary time series with logical covariance $\tau_{\epsilon\epsilon}(j)$ and

Walsh-Fourier spectrum $f_{\epsilon\epsilon}(\lambda)$. As in the model (3.3), $S_\ell(n)$ and $\epsilon_{q\ell}(n)$ are uncorrelated, $q = 1, \dots, Q$, however $S_j(n)$ and $\epsilon_{q\ell}(n)$ are mutually independent for $j \neq \ell$ and $q = 1, \dots, Q$. We assume that every linear combination of $S_\ell(n)$ and $\epsilon_{q\ell}(n)$, $q = 1, \dots, Q$, satisfies either Assumption 2.1, 2.2 or 2.3, $\ell = 1, \dots, L$. Under the model assumptions we have

- (1) $EX_{q\ell}^{(n)} = \theta_{q\ell}$,
- (2) $\tau_{X_{q\ell}, X_{q\ell}}^{(j)} = \tau_{SS}^{(j)} + \tau_{\epsilon\epsilon}^{(j)}$,
- (3) $f_{X_{q\ell}, X_{q\ell}}^{(\lambda)} = f_{SS}^{(\lambda)} + f_{\epsilon\epsilon}^{(\lambda)}$,
- (4) $\tau_{X_{q\ell}, X_{p\ell}}^{(j)} = \tau_{SS}^{(j)}, \quad q \neq p$
- (5) $f_{X_{q\ell}, X_{p\ell}}^{(\lambda)} = f_{SS}^{(\lambda)}, \quad q \neq p.$

It is seen that the coherence between any two observations in the same treatment is

$$f_{SS}^{(\lambda)} / [f_{SS}^{(\lambda)} + f_{\epsilon\epsilon}^{(\lambda)}].$$

This corresponds to the intraclass correlation coefficient discussed in Scheffé (1959).

Taking Walsh-Fourier transforms in (3.11) and following Theorem 3.1 we have the representation

$$d_{N, q\ell}^{(\lambda)} j(m), N = u_{\ell m} + z_{q\ell m} \quad \text{a.s.} \quad (3.12)$$

as $N \rightarrow \infty$, $m = 1, \dots, M$, where the $u_{\ell m}$ are independent $N(0, f_{SS}^{(\lambda)})$ variates and the $z_{q\ell m}$ are independent $N(0, f_{\epsilon\epsilon}^{(\lambda)})$ variates which are independent of

the $u_{\ell m}$, all q , ℓ and m . Using the representation (3.12) we consider a test for determining whether or not there is a treatment effect, that is, whether or not $S_{\ell}(n) \equiv 0$ for $\ell = 1, \dots, L$, and all n . Equivalently, we test whether or not $f_{SS}(\lambda) \equiv 0$, $0 < \lambda < 1$. We may now follow the one-way random effects analysis given in Scheffé (1969, Section 7.2). Define

$$d_{N, \cdot \ell}(\lambda_{j(m), N}) = Q^{-1} \sum_{q=1}^Q d_{N, q\ell}(\lambda_{j(m), N}), \quad 1 \leq m \leq M$$

$$d_{N, \cdot \cdot}(\lambda_{j(m), N}) = L^{-1} \sum_{\ell=1}^L d_{N, \cdot \ell}(\lambda_{j(m), N}), \quad 1 \leq m \leq M$$

and appropriate sums of squares

$$Q \sum_{m=1}^M \sum_{\ell=1}^L [d_{N, \cdot \ell}(\lambda_{j(m), N}) - d_{N, \cdot \cdot}(\lambda_{j(m), N})]^2 \quad (3.13)$$

$$\sum_{m=1}^M \sum_{\ell=1}^L \sum_{q=1}^Q [d_{N, q\ell}(\lambda_{j(m), N}) - d_{N, \cdot \ell}(\lambda_{j(m), N})]^2. \quad (3.14)$$

In view of (3.12), (3.13) and (3.14) have the almost sure representations

$$Q \sum_{m=1}^M \sum_{\ell=1}^L (u_{\ell m} + z_{\cdot \ell m} - u_{\cdot m} - z_{\cdot \cdot m})^2 \quad (3.15)$$

and

$$\sum_{m=1}^M \sum_{\ell=1}^L \sum_{q=1}^Q (z_{q\ell m} - z_{\cdot \ell m})^2 \quad (3.16)$$

respectively, as $N \rightarrow \infty$. We may carry this representation one step further and

represent (3.15) and (3.16) as

$$[Q f_{SS}(\lambda) + f_{\epsilon\epsilon}(\lambda)] \chi_{M(L-1)}^2 \quad \text{a.s.} \quad (3.17)$$

and

$$f_{\epsilon\epsilon}(\lambda) \chi_{ML(Q-1)}^2 \quad \text{a.s.} \quad (3.18)$$

as $N \rightarrow \infty$, where the χ^2 variates in (3.17) and (3.18) are independent.

The test of the null hypothesis $f_{SS}(\lambda) \equiv 0$ now follows easily in view of (3.17) and (3.18). For a particular sequency λ , compare the statistic

$$\frac{Q \sum_{m=1}^M \sum_{\ell=1}^L [d_{N,\ell}(\lambda_{j(m),N}) - d_{N,\cdot}(\lambda_{j(m),N})]^2 / M(L-1)}{\sum_{m=1}^M \sum_{\ell=1}^L \sum_{q=1}^Q [d_{N,q}(\lambda_{j(m),N}) - d_{N,\ell}(\lambda_{j(m),N})]^2 / ML(Q-1)}$$

with an $F_{M(L-1), ML(Q-1)}$ distribution.

To estimate the Walsh-Fourier spectrum of the signal, $f_{SS}(\lambda)$, one may use representations (3.15), (3.16), (3.17) and (3.18). That is, the expected value of (3.15) divided by $M(L-1)$ (cf. 3.17) is $Q f_{SS}(\lambda) + f_{\epsilon\epsilon}(\lambda)$ and the expected value of (3.16) divided by $ML(Q-1)$ (cf. 3.18) is $f_{\epsilon\epsilon}(\lambda)$. This is similar to the previous case (cf. 3.10) and one may proceed as was done there.

As suggested in Brillinger (1973, 1980), it is helpful to graph the above statistic as a function of sequency, λ , and to further indicate the null significance on the graph. However, in doing so, one must keep in mind that the asymptotic ($N \rightarrow \infty$) covariance of the Walsh-Fourier transform at two distinct frequencies is not necessarily zero in contrast to the trigonometric case.

3.2 Estimating Regression Parameters When the Discrete Signal is Observable

In this section we concentrate on the discrete signal-plus-noise model where the signal is observable but the vector process of interest, $\underline{X}(n)$, $n = 0, 1, \dots, N-1$, depends on unknown regression parameters. At the end of this section we apply the results of this section to estimating the transition matrix in a Markov model. We suppose that we may write the model (3.1) in the $r \times 1$ vector form

$$\underline{X}(n) = \underline{\theta} + \beta \underline{S}(n) + \underline{\varepsilon}(n), \quad (3.19)$$

where $\underline{\theta}$ is an $r \times 1$ vector of constants, $\underline{S}(n)$ is a $q \times 1$ vector, observable discrete-valued signal, β is an $r \times q$ matrix of regression parameters, and $\underline{\varepsilon}(n)$ is a discrete-valued zero-mean white noise process which is uncorrelated with the signal $\underline{S}(n)$. Let $\Gamma_{SS}(h)$ denote the autocovariance matrix of $\underline{S}(n)$, where $\sum_{-\infty}^{\infty} \|\Gamma_{SS}(h)\| < \infty$. Further, we assume that every linear combination of the components of $\underline{\varepsilon}(n)$ satisfy either Assumption 2.1, 2.2 or 2.3. Let $\tau_{SS}(j)$ and $\tau_{\varepsilon\varepsilon}(j)$ be the logical covariance matrices of $\underline{S}(n)$ and $\underline{\varepsilon}(n)$, respectively (cf. 2.9) and note that $\tau_{\varepsilon\varepsilon}(j) = 0_r$ for $j \neq 0$ where 0_r is the $r \times r$ matrix of zeroes. Let $f_{SS}(\lambda) = \{f_{S_a, S_b}(\lambda)\}$, $1 \leq a, b \leq q$, be the $q \times q$ positive definite Walsh-Fourier spectrum of $\underline{S}(n)$ and let $f_{\varepsilon\varepsilon}(\lambda) = \{f_{\varepsilon_a, \varepsilon_b}(\lambda)\}$, $1 \leq a, b \leq r$, be the $r \times r$ Walsh-Fourier spectrum of $\underline{\varepsilon}(n)$.

Taking transforms in (3.19) and following the previous section we have the representation

$$\underline{d}_N^X(\lambda_{j(m), N}) = \beta \underline{d}_N^S(\lambda_{j(m), N}) + \underline{z}_m + \underline{0}_{a.s.}(1), \quad m = 1, \dots, M, \quad (3.20)$$

where $Q_{a.s.}^{(1)}$ denotes a vector variate tending to the zero vector almost surely as $N \rightarrow \infty$, z_m are independent multivariate $N(Q, f_{\epsilon\epsilon}(\lambda))$ variates:

$d_N^{X(\lambda_{j(m),N})} = N^{-1/2} \sum_{n=0}^{N-1} X(n)W(n, \lambda_{j(m),N})$ and $d_N^{S(\lambda_{j(m),N})} = N^{-1/2} \sum_{n=0}^{N-1} S(n)W(n, \lambda_{j(m),N})$ denote the $r \times 1$ and $q \times 1$ finite Walsh-Fourier transforms of the vector time series $X(n)$ and $S(n)$, respectively.

If we assume that $M \geq q+r$, we may consider the asymptotic problem of estimating the parameter matrix β and the error spectrum in the multivariate analysis setting (see Anderson (1958), Chapter 8). Similarly, tests of hypotheses about the elements of β may be carried out by MANOVA techniques with an appropriate partitioning of that matrix (cf. Anderson (1958), Section 8.3). Following least squares theory in view of (3.20), the estimate of β as a function of frequency is given by

$$\hat{\beta}_{M,N}(\lambda) = \left[\sum_{m=1}^M d_N^{X(\lambda_{j(m),N})} d_N^{S'(\lambda_{j(m),N})} \right] \left[\sum_{m=1}^M d_N^{S(\lambda_{j(m),N})} d_N^{S'(\lambda_{j(m),N})} \right]^{-1} \quad (3.21)$$

and the corresponding estimate of the error spectrum $f_{\epsilon\epsilon}(\lambda)$ is

$$\begin{aligned} \hat{f}_{\epsilon,\epsilon}^{M,N}(\lambda) = M^{-1} \sum_{m=1}^M & \left[d_N^{X(\lambda_{j(m),N})} - \hat{\beta}_{M,N}(\lambda) d_N^{S(\lambda_{j(m),N})} \right] \left[d_N^{X(\lambda_{j(m),N})} \right. \\ & \left. - \hat{\beta}_{M,N}(\lambda) d_N^{S(\lambda_{j(m),N})} \right]' . \end{aligned} \quad (3.22)$$

Before discussing the consistency of the estimates $\hat{\beta}_{M,N}(\lambda)$ and $\hat{f}_{\epsilon,\epsilon}^{M,N}(\lambda)$, we state the following results which follow from the model assumptions and the results of Section 2. First, denote

$$f_{S_a, S_b}^{M, N}(\lambda) = M^{-1} \sum_{m=1}^M d_{N,a}^S(\lambda_{j(m), N}) d_{N,b}^S(\lambda_{j(m), N}), \quad 1 \leq a, b \leq q \quad (3.23 \text{ a})$$

$$f_{\epsilon_a, S_b}^{M, N}(\lambda) = M^{-1} \sum_{m=1}^M d_{N,a}^\epsilon(\lambda_{j(m), N}) d_{N,b}^S(\lambda_{j(m), N}), \quad 1 \leq a \leq r, \quad 1 \leq b \leq q \quad (3.23 \text{ b})$$

and

$$f_{\epsilon_a, \epsilon_b}^{M, N}(\lambda) = M^{-1} \sum_{m=1}^M d_{N,a}^\epsilon(\lambda_{j(m), N}) d_{N,b}^\epsilon(\lambda_{j(m), N}), \quad 1 \leq a, b \leq r \quad (3.23 \text{ c})$$

where $d_{N,a}^S(\cdot)$ and $d_{N,a}^\epsilon(\cdot)$ denote the finite Walsh-Fourier transform of the a -th element of $\xi(n)$ and $\epsilon(n)$, respectively. Then as $M, N \rightarrow \infty$ with $M/N \rightarrow 0$ we have

$$(1) \quad E f_{S_a, S_b}^{M, N}(\lambda) \rightarrow f_{S_a, S_b}(\lambda),$$

$$(2) \quad E f_{\epsilon_a, \epsilon_b}^{M, N}(\lambda) \rightarrow f_{\epsilon_a, \epsilon_b}(\lambda),$$

$$(3) \quad E f_{\epsilon_a, S_b}^{M, N}(\lambda) \equiv 0,$$

$$(4) \quad M \text{Var}\{f_{S_a, S_b}^{M, N}(\lambda)\} \rightarrow 2[f_{S_a, S_b}(\lambda)]^2,$$

$$(5) \quad M \text{Var}\{f_{\epsilon_a, \epsilon_b}^{M, N}(\lambda)\} \rightarrow 2[f_{\epsilon_a, \epsilon_b}(\lambda)]^2,$$

$$(6) \quad M \text{Var}\{f_{\epsilon_a, S_b}^{M, N}(\lambda)\} \rightarrow f_{\epsilon_a, \epsilon_a}(\lambda) f_{S_b, S_b}(\lambda).$$

It follows from the above that as $M, N \rightarrow \infty$ with $M/N \rightarrow 0$, $f_{S_a, S_b}^{M, N}(\lambda) \rightarrow f_{S_a, S_b}(\lambda)$,

$f_{\epsilon_a, \epsilon_b}^{M,N}(\lambda) \rightarrow f_{\epsilon_a, \epsilon_b}(\lambda)$, and $f_{\epsilon_a, S_b}^{M,N}(\lambda) \rightarrow 0$ in mean square and hence in probability.

We now show the consistency of the estimates (3.21) and (3.22).

THEOREM 3.2: Let the model assumptions (3.19) be satisfied. Then as $M, N \rightarrow \infty$ with $M/N \rightarrow 0$ the estimates $\hat{\beta}_{M,N}(\lambda)$ given in (3.21) and $\hat{f}_{\epsilon\epsilon}^{M,N}(\lambda)$ given in (3.22) are consistent in probability for β and $f_{\epsilon\epsilon}(\lambda)$, $0 < \lambda < 1$, respectively.

PROOF: Using the definitions of (3.23), write

$$\hat{\beta}_{M,N}(\lambda) = \beta + f_{\epsilon, S}^{M,N}(\lambda) [f_{SS}^{M,N}(\lambda)]^{-1} \quad (3.24)$$

where $f_{\epsilon, S}^{M,N}(\lambda) = \{f_{\epsilon_a, S_b}^{M,N}(\lambda)\}$ and $f_{SS}^{M,N}(\lambda) = \{f_{S_a, S_b}^{M,N}(\lambda)\}$ are $r \times q$ and $q \times q$ matrices, respectively. By previous results we know that as $M, N \rightarrow \infty$ with $M/N \rightarrow 0$, $f_{SS}^{M,N}(\lambda) \xrightarrow{P} f_{SS}(\lambda)$ which is positive definite, and $f_{\epsilon, S}^{M,N}(\lambda) \xrightarrow{P} 0_{r \times q}$ where $0_{r \times q}$ is an $r \times q$ matrix of zeroes. In view of (3.24) we see that $\hat{\beta}_{M,N}(\lambda)$ is consistent for β .

To show the consistency of $\hat{f}_{\epsilon\epsilon}^{M,N}(\lambda)$ given by (3.22), first note that we may write the $r \times r$ matrix $f_{\epsilon\epsilon}^{M,N}(\lambda) = \{f_{\epsilon_a, \epsilon_b}^{M,N}(\lambda)\}$ whose elements are given by (3.23 c) as

$$f_{\epsilon\epsilon}^{M,N}(\lambda) = M^{-1} \sum_{m=1}^M [d_N^X(\lambda_{j(m), N}) - \beta d_N^S(\lambda_{j(m), N})][d_N^X(\lambda_{j(m), N}) - \beta d_N^S(\lambda_{j(m), N})]'$$

Since $f_{\epsilon\epsilon}^{M,N}(\lambda)$ is consistent for $f_{\epsilon\epsilon}(\lambda)$, it suffices to show that

$\hat{f}_{\epsilon\epsilon}^{M,N}(\lambda) - f_{\epsilon\epsilon}^{M,N}(\lambda) \rightarrow 0$ in probability as $M, N \rightarrow \infty$ with $M/N \rightarrow 0$. To see this

write

$$\begin{aligned}
 \hat{f}_{\varepsilon\varepsilon}^{M,N}(\lambda) - f_{\varepsilon\varepsilon}^{M,N}(\lambda) &= M^{-1} \sum_{m=1}^M d_N^X(\lambda_{j(m),N}) d_N^{S'}(\lambda_{j(m),N}) [\beta - \hat{\beta}_{M,N}(\lambda)]' \\
 &+ [\beta - \hat{\beta}_{M,N}(\lambda)] M^{-1} \sum_{m=1}^M d_N^S(\lambda_{j(m),N}) d_N^{X'}(\lambda_{j(m),N}) \\
 &+ [\hat{\beta}_{M,N}(\lambda) - \beta] M^{-1} \sum_{m=1}^M d_N^S(\lambda_{j(m),N}) d_N^{S'}(\lambda_{j(m),N}) \beta' \\
 &+ \hat{\beta}_{M,N}(\lambda) M^{-1} \sum_{m=1}^M d_N^S(\lambda_{j(m),N}) d_N^{S'}(\lambda_{j(m),N}) [\hat{\beta}_{M,N}(\lambda) - \beta]. \quad (3.25)
 \end{aligned}$$

We consider the first term of (3.25), the other cases will follow by similar methods. Expanding $d_N^X(\lambda_{j(m),N})$ we have

$$\begin{aligned}
 M^{-1} \sum_{m=1}^M d_N^X(\lambda_{j(m),N}) d_N^{S'}(\lambda_{j(m),N}) [\beta - \hat{\beta}_{M,N}(\lambda)]' &= \beta f_{SS}^{M,N}(\lambda) [\beta - \hat{\beta}_{M,N}(\lambda)]' \\
 &+ f_{\varepsilon S}^{M,N}(\lambda) [\beta - \hat{\beta}_{M,N}(\lambda)]' \rightarrow 0_r
 \end{aligned}$$

in probability as $M, N \rightarrow \infty$ with $M/N \rightarrow 0$ by (3.24) and the convergence results for $f_{SS}^{M,N}(\lambda)$ and $f_{\varepsilon S}^{M,N}(\lambda)$.

We close this section with an application to Markov chains. In particular, consider the macro model given by (3.2) and suppose that we are interested in estimating the transition probabilities p_{ij} . We shall assume that detailed transitions are not available, otherwise, the p_{ij} may be estimated by MLE techniques, see Basawa and Prakasa Rao (1980, Section 2.1). Recall that the

model is of the form

$$X_\ell(n) = \sum_{k=1}^r X_k(n-1)p_{k\ell} + \epsilon_\ell(n), \quad 1 \leq \ell \leq r, \quad (3.26)$$

where $X_\ell(n)$ is the (aggregate) number of individuals in state ℓ (over several independent Markov chains) at time n and $\epsilon_\ell(n)$ is zero-mean white noise which is uncorrelated with $X_k(n-1)$ for $k = 1, \dots, r$, and r is the number of states. We suppose that data is available in the form $\{X_\ell(n), \ell = 1, \dots, r; n = 0, \dots, N-1\}$ and that the total number of individuals in the system, viz. $\sum_{\ell=1}^r X_\ell(n)$, is fixed. To avoid singularities we remove $X_r(n)$ from the model (3.26). Let T be the total number of individuals in the system, then for $\ell = 1, \dots, r-1$,

$$X_\ell(n) = T(1 - \sum_{j=1}^{r-1} p_{j\ell}) + \sum_{j=1}^{r-1} [X_j(n-1) + \sum_{k=1}^{r-1} X_k(n-1)]p_{j\ell} + \epsilon_\ell(n). \quad (3.27)$$

Now, put $\theta_\ell = T(1 - \sum_{j=1}^{r-1} p_{j\ell})$, $S_j(n) = X_j(n-1) + \sum_{k=1}^{r-1} X_k(n-1)$ with $S_j(0) \equiv 0$, and let $\underline{X}(n) = (X_1(n), \dots, X_{r-1}(n))'$, $\underline{\theta} = (\theta_1, \dots, \theta_{r-1})'$, $\underline{S}(n) = (S_1(n), \dots, S_{r-1}(n))'$ and $\underline{\epsilon}(n) = (\epsilon_1(n), \dots, \epsilon_{r-1}(n))'$. We have already seen that $\underline{\epsilon}(n)$ and $\underline{S}(n)$ are uncorrelated. In view of (3.27) the model (3.26) may now be represented in the form of (3.19), that is

$$\underline{X}(n) = \underline{\theta} + P \underline{S}(n) + \underline{\epsilon}(n), \quad n \geq 1, \quad (3.28)$$

where P is the $(r-1) \times (r-1)$ matrix of unknown transition probabilities

$$P = \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{r-1,1} \\ p_{12} & p_{22} & \cdots & p_{r-1,2} \\ \vdots & \vdots & & \vdots \\ p_{1,r-1} & p_{2,r-1} & \cdots & p_{r-1,r-1} \end{bmatrix}.$$

If we choose M frequencies, $\lambda_{j(m),N}$, $m = 1, \dots, M$, $\lambda_{j(m),N} \neq 0$ in such a way that $M \geq 2(r-1)$ we obtain the consistent estimate of the transition matrix

$$\hat{P}(\lambda) = f_{XS}^{M,N}(\lambda) [f_{SS}^{M,N}(\lambda)]^{-1},$$

$$\text{where } f_{XS}^{M,N}(\lambda) = M^{-1} \sum_{m=1}^M d_N^X(\lambda_{j(m),N}) d_N^{S'}(\lambda_{j(m),N}),$$

$$f_{SS}^{M,N}(\lambda) = M^{-1} \sum_{m=1}^M d_N^S(\lambda_{j(m),N}) d_N^{S'}(\lambda_{j(m),N}),$$

$$d_N^X(\lambda_{j(m),N}) = N^{-1/2} \sum_{n=0}^{N-1} X(n) W(n, \lambda_{j(m),N}) \quad \text{and}$$

$$d_N^S(\lambda_{j(m),N}) = N^{-1/2} \sum_{n=1}^{N-1} \{X(n-1) + \sum_{k=1}^{r-1} X(n-1)\} W(n, \lambda_{j(m),N}).$$

In this case we also obtain a consistent estimate for θ , namely,

$$\hat{\theta}_\ell(\lambda) = T(1 - \sum_{j=1}^{r-1} \hat{p}_{j\ell}(\lambda)), \quad \ell = 1, \dots, r-1,$$

where $\hat{p}_{j\ell}(\lambda)$ is the $j\ell$ -th element of $\hat{P}(\lambda)$. The consistent estimate of the error spectrum is

$$\hat{f}_{\epsilon, \epsilon}(\lambda) = f_{XX}^{M,N}(\lambda) - \hat{P}(\lambda) f_{SS}^{M,N} \hat{P}'(\lambda).$$

The actual implementation of this estimation procedure is then based on the methods described in the Appendix A.

APPENDIX A

In this appendix we discuss the computation of the finite Walsh-Fourier transform for samples of length $N = 2^p$, $p > 0$ integer. The discrete Walsh functions are calculated via the Hadamard matrix, $H(p)$, which is defined to be the symmetric orthogonal $N \times N$ matrix whose (u,v) -th element, $u, v = 0, 1, \dots, N-1$, is equal to $\phi(u,v) = \prod_{j=0}^{p-1} (-1)^{u_j v_j}$ where the binary representations of u and v are given by $(u_{p-1}, u_{p-2}, \dots, u_0)$ and $(v_{p-1}, v_{p-2}, \dots, v_0)$, respectively, u_j and v_j are either 0 or 1. For example, with $N = 8$, the $(1,5)$ element of $H(3)$ is $\phi(1,5) = (-1)^{0+0+1} = -1$ whereas the $(3,7)$ element is $\phi(3,7) = (-1)^{0+1+1} = 1$. The Hadamard matrix gives the discrete Walsh functions as rows (or columns) in what is called natural or Hadamard ordering. To obtain the Walsh functions in sequency order, we can reorder the rows of $H(p)$ according to the number of sign changes. We denote the sequency or Walsh ordered Hadamard matrix by $H_W(p)$. An alternate method of obtaining $H_W(p)$ from $H(p)$ uses "bit-reversal Gray code" to rearrange the rows, however this is essentially the same technique as counting the sign changes, see Ahmed and Rao (1975) for details. Since these approaches involve counting sign changes, they are not very efficient procedures. We shall discuss a fast Walsh-ordered Hadamard transform in a moment. The Hadamard matrix can be generated recursively, $H(0) = 1$, and

$$H(k+1) = \begin{bmatrix} H(k) & H(k) \\ H(k) & -H(k) \end{bmatrix}, \quad k = 0, 1, 2, \dots$$

so for example

$$H(1) = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad \text{and} \quad H(2) = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}.$$

Denote $H(2) = [h_0(2), h_1(2), h_2(2), h_3(2)]$ where $h_i(2)$, $i = 0, 1, 2, 3$ is the i -th column of $H(2)$, then the corresponding Walsh-ordered Hadamard matrix is $H_W(2) = [h_0(2), h_2(2), h_3(2), h_1(2)]$. The procedure of obtaining the Walsh-ordered Hadamard transform from its definition either requires storage of the Hadamard matrix, or recomputation whenever the elements of $H_W(p)$ are needed. Hence, either the sample length is restricted to about $p = 10$ or 20 , or the procedure is extremely slow. There are, however, fast methods which can reduce the number of computations (additions and subtractions) by about $2^{p-1}/p$ times less than by using the definition. The method we discuss here may be found in Ahmed and Rao (1975, Chapter 6) who also gives a computer subroutine. The Walsh-Hadamard matrix can be computed as

$$H_W(p) = \prod_{i=1}^p H_i(p) \cdot B \quad (A1)$$

where

$$H_i(p) = \begin{bmatrix} F_s & G_s & & & & & 0 \\ & & \ddots & & & & \\ & & & F_s & G_s & & \\ & & & & & \ddots & \\ & & & & & & F_s & G_s \\ 0 & & & & & & & \end{bmatrix}, \quad s = 2^{i-1}$$

with

$$F_s = \begin{bmatrix} I_s & I_s \\ I_s & -I_s \end{bmatrix}$$

$$G_s = \begin{bmatrix} I_s & -I_s \\ I_s & I_s \end{bmatrix}$$

and I_s being the $s \times s$ identity matrix. The matrix B in (A1) is a matrix which bit-reverses the order of the data. For example, with $N = 2^3$ the bit-reversal of $1 = (0,0,1)$ is $4 = (1,0,0)$, and the bit-reversal of $3 = (0,1,1)$ is $6 = (1,1,0)$ so that $X(1)$ is exchanged with $X(4)$ and $X(3)$ is exchanged with $X(6)$ in the data vector. If $\underline{X} = (X(0), \dots, X(N-1))'$ is the data vector, $N = 2^p$, the fast finite Walsh-Fourier transform is computed as

$$d_N(\lambda_N) = N^{-1/2} H_W(p) \underline{X} = N^{-1/2} \prod_{i=1}^p H_i(p) \cdot B \underline{X} \quad (A2)$$

where $\lambda_N = (0/N, 1/N, \dots, (N-1)/N)'$. For example, if $N = 2^3$, the Walsh-ordered Hadamard matrix can be decomposed as

$$H_W(3) = \begin{bmatrix} F_1 & & & 0 \\ & G_1 & & \\ & & F_1 & \\ 0 & & & G_1 \end{bmatrix} \begin{bmatrix} F_2 & 0 \\ & G_2 \\ 0 & \\ & \end{bmatrix} F_3 B$$

where

$$B = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

The vector of periodogram ordinates is obtained by squaring each element of $d_N(\lambda_N)$ given in (A2). Let $I_N(m) = d_N^2(m/N)$ denote the m -th periodogram ordinate, $m = 0, 1, \dots, N-1$. It is seen that $I_N(m) = \sum_{j=0}^{N-1} \hat{\tau}(j) W(j, m/N)$ where $\hat{\tau}(j) = N^{-1} \sum_{k=0}^{N-1} X(k) X(j \oplus k)$. Employing relationships (2.5) and (2.6) we may write

$$\hat{\tau}(j) = N^{-1} \sum_{m=0}^{N-1} I_N(m) W(m, j/N). \quad (A3)$$

Thus, for large N , the quickest way to compute $\hat{\tau}(j)$ is to use the fast Walsh-Fourier transform twice, once to compute $I_N(m)$ and once to compute the right hand side of (A3).

APPENDIX B

In this appendix we prove the central limit theorem stated in Section 2 which is based on condition (2.4) and Assumption 2.3. The proof of the theorem is based on a martingale central limit theorem given in Brown (1971, Theorem 2) and a limit theorem for approximating sequences given in Anderson (1971, Theorem 7.7.1). We now state and prove the theorem.

THEOREM B: Let $\{X(n); n = 0, \pm 1, \pm 2, \dots\}$ satisfy condition (2.4) and Assumption 2.3. Then the discrete Walsh-Fourier transform, $d_N(\lambda)$, converges in distribution to a normal variate with mean zero and variance $f(\lambda)$ given by (2.3).

PROOF: Fix $\lambda \in [0, 1)$ and write $d_N(\lambda) = N^{-1/2} \sum_{n=0}^{N-1} X(n)W(n, \lambda) \equiv N^{-1/2} \sum_{n=1}^N y_n$. Choose an integer k such that $2k < k < N$ and put $T = [N/k]$ where $[\cdot]$ is the greatest integer function. For $j = 1, 2, \dots, T$ define $Y_{j,k} = y_{(j-1)k+1} + y_{(j-1)k+2} + \dots + y_{jk-k}$ and $Z_{j,k} = y_{jk-k+1} + y_{jk-k+2} + \dots + y_{jk}$, and let $u_{k,N} = N^{-1/2} \sum_{j=1}^T Z_{j,k}$ and $v_{k,N} = N^{-1/2} (y_{kT+1} + y_{kT+2} + \dots + y_N)$. Clearly $d_N(\lambda) = N^{-1/2} \sum_{j=1}^T Y_{j,k} + u_{k,N} + v_{k,N}$. We will show that $v_{k,N} \rightarrow 0$ in mean square as $N \rightarrow \infty$ uniformly in k , $u_{k,N} \rightarrow 0$ in mean square as $k \rightarrow \infty$ uniformly in N , and $N^{-1/2} \sum_{j=1}^T Y_{j,k}$ converges in law to a normal variate as $N \rightarrow \infty$ and $k \rightarrow \infty$. Thus, the theorem will follow from Anderson (1971, Theorem 7.7.1).

Using Minkowski's inequality with $r = N - Tk < k$ and the fact that

$$|W(n, \lambda)| = 1,$$

$$\begin{aligned}
E\{v_{k,N}^2\}^{1/2} &= N^{-1/2} \{E(y_{kT+1} + \dots + y_N)^2\}^{1/2} \\
&\leq N^{-1/2} \sum_{j=1}^r E\{|X(kT+j)|^2\}^{1/2}
\end{aligned}$$

from which we obtain

$$E\{v_{k,N}^2\} \leq r^2 N^{-1} \gamma(0) \leq k^2 N^{-1} \gamma(0)$$

and hence the desired result follows. In a similar manner we can show that

$$E\{u_{k,N}^2\} \leq T N^{-1} \kappa^2 \gamma(0) \leq k^{-1} \kappa^2 \gamma(0)$$

to obtain the desired result for $u_{k,N}$.

To establish a central limit theorem for $N^{-1/2} \sum_{j=1}^T Y_{j,k}$ we use a martingale central limit theorem found in Brown (1971, Theorem 2). We shall follow Brown's notation. Let k be fixed and let $S_{k,T} = \sum_{j=1}^T Y_{j,k}$; recall that F_n is the σ -field generated by $\{X(j), j \leq n\}$. Then, $\{S_{k,T}, F_{kT-\kappa}; T = 1, 2, \dots\}$ form a martingale. To see this note that

$$E\{S_{k,T+1} \mid F_{kT-\kappa}\} = \sum_{j=1}^T Y_{j,k} + E\{Y_{T+1,k} \mid F_{kT-\kappa}\}.$$

The terms in $Y_{T+1,k}$ are $\{X(n)W(n,\lambda), n = kT + j; j = 1, 2, \dots, k-\kappa\}$. We are given that $E\{X(kT+j) \mid F_{kT+j-\kappa}\} = 0$ a.s. and since $F_{kT-\kappa} \subset F_{kT+j-\kappa}$, $1 \leq j \leq k-\kappa$, we have

$$E\{X(kT+j) \mid F_{kT-\kappa}\} = E\{E\{X(kT+j) \mid F_{kT+j-\kappa}\} \mid F_{kT-\kappa}\} = 0 \quad \text{a.s.}$$

and hence $E\{Y_{T+1,k} | F_{kT-\kappa}\} = 0$ a.s. as was to be shown. Following Brown (1971), put $V_{k,T}^2 = \sum_{j=1}^T E\{Y_j^2 | F_{k(j-1)-\kappa}\}$ and $s_{k,T}^2 = EV_{k,T}^2$. A typical term in $V_{k,T}^2$ is $E\{Y_{(j-1)k+\ell} Y_{(j-1)k+m} | F_{k(j-1)-\kappa}\}$, $1 \leq \ell, m \leq k-\kappa$; $1 \leq j \leq T$. Let $n_1 = (j-1)k + \ell$ and $n_2 = (j-1)k + m$, then by hypotheses of the theorem,

$$\begin{aligned} E\{Y_{n_1} Y_{n_2} | F_{(j-1)k-\kappa}\} &= W(n_1 \oplus n_2, \lambda) E\{X(n_1) X(n_2) | F_{(j-1)k-\kappa}\} \\ &= W(n_1 \oplus n_2, \lambda) \gamma(n_2 - n_1) \quad \text{a.s.} \\ &= E\{Y_{n_1} Y_{n_2}\}. \end{aligned}$$

Hence, $V_{k,T}^2 = \sum_{j=1}^T EY_j^2 = EV_{k,T}^2 = s_{k,T}^2$ a.s. The application of Theorem 2 of Brown (1977) entails showing that as $T \rightarrow \infty$,

$$\begin{aligned} \text{(i)} \quad & s_{k,T}^{-2} V_{k,T}^2 \xrightarrow{p} 1, \text{ and} \\ \text{(ii)} \quad & s_{k,T}^{-2} \sum_{j=1}^T E\{Y_{j,k}^2 I(|Y_{j,k}| \geq \varepsilon s_{k,T})\} \rightarrow 0 \end{aligned}$$

for any $\varepsilon > 0$, where $I(\cdot)$ is the indicator function. We have already established (i) and we proceed with establishing (ii). First, we show that $\sup_n E\{|X(n)|^{2+\delta}\} < \infty$ for some $\delta > 0$ implies that $\sup_j E\{|Y_{j,k}|^{2+\delta}\} < \infty$. Let $n_\ell = (j-1)k + \ell$ and $a(n_\ell) = W(n_\ell, \lambda)$. Then by Minkowski's inequality

$$\begin{aligned} E\{|Y_{j,k}|^{2+\delta}\}^{\frac{1}{2+\delta}} &= E\{|a(n_1)X(n_1) + \dots + a(n_{k-\kappa})X(n_{k-\kappa})|^{2+\delta}\}^{\frac{1}{2+\delta}} \\ &\leq \sum_{\ell=1}^{k-\kappa} E\{|X(n_\ell)|^{2+\delta}\}^{\frac{1}{2+\delta}} \\ &\leq (k-\kappa) [\sup_n E\{|X(n)|^{2+\delta}\}]^{\frac{1}{2+\delta}} \end{aligned}$$

so that for any j ,

$$E\{|Y_{j,k}|^{2+\delta}\} \leq (k-\kappa)^{2+\delta} \sup_n E\{|X(n)|^{2+\delta}\} < \infty$$

as was to be shown. Following a similar argument to the one above, we can show that $EY_{j,k}^2$ is uniformly bounded, and since $EY_{j,k}^2$ does not vanish as $j \rightarrow \infty$ we see that $s_{k,T}^2 = O(T)$. Also, one can show that since $\sup_j E\{|Y_{j,k}|^{2+\delta}\} < \infty$ for some $\delta > 0$, there exists a random variable Y_k with $EY_k^2 < \infty$ such that

$$E\{Y_{j,k}^2 I(|Y_{j,k}| \geq u)\} \leq E\{Y_k^2 I(|Y_k| \geq u)\}$$

for all j and all $u \geq 0$. See Stoffer (1985, Lemma 3.2). Thus,

$$s_{k,T}^{-2} \sum_{j=1}^T E\{Y_{j,k}^2 I(|Y_{j,k}| \geq \epsilon s_{k,T})\} \leq s_{k,T}^{-2} E\{Y_k^2 I(|Y_k| \geq \epsilon s_{k,T})\} \rightarrow 0$$

as $T \rightarrow \infty$. Hence by the Brown (1971) martingale central limit theorem we have that for fixed k , $T^{-1/2} \sum_{j=1}^T Y_{j,k}^2 = (N^{1/2} T^{-1/2}) N^{-1/2} \sum_{j=1}^T Y_{j,k}^2$ converges in law to a normal variate whose variance depends on k as $T \rightarrow \infty$ or equivalently $N \rightarrow \infty$. Clearly then, for fixed k , $(N^{1/2} T^{-1/2} k^{-1/2}) N^{-1/2} \sum_{j=1}^T Y_{j,k}^2$ converges in law to a normal variate as $N \rightarrow \infty$. Letting $k \rightarrow \infty$ establishes the desired result.

REFERENCES

- [1] Ahmed, N. and Rao, K. R. (1975). Orthogonal Transforms for Digital Signal Processing, Springer-Verlag, New York.
- [2] Anderson, T. W. (1958). An Introduction to Multivariate Statistical Analysis, Wiley, New York.
- [3] Anderson, T. W. (1971). The Statistical Analysis of Time Series, Wiley and Sons, New York.
- [4] Basawa, I. V. and Prakasa Rao, B. L. S. (1980). Statistical Inference for Stochastic Processes, Academic Press, London.
- [5] Beauchamp, K. G. (1975). Walsh Functions and Their Applications, Academic Press, London.
- [6] Brillinger, D. R. (1973). The Analysis of Time Series Collected in an Experimental Design, Multivariate Analysis III, ed. by P. R. Krishnaiah, Academic Press, New York, 241-256.
- [7] Brillinger, D. R. (1975). Time Series: Data Analysis and Theory, Holt, Rinehart and Winston, Inc., New York.
- [8] Brillinger, D. R. (1980). Analysis of Variance Problems Under Time Series Models, Handbook of Statistics, Vol. 1, ed. by P. R. Krishnaiah, North Holland, 237-278.
- [9] Brillinger, D. R. and Tukey, J. W. (1982). Spectrum Estimation and System Identification Relying on a Fourier Transform, Technical Report No. 5, Department of Statistics, Berkeley, California.
- [10] Brown, B. M. (1971). Martingale Central Limit Theorems, Ann. Math. Statist. 42, 59-66.
- [11] Chung, K. L. (1974). A Course in Probability Theory, Academic Press, New York.
- [12] Hannan, E. J. (1973). Central Limit Theorems for Time Series Regression, Z. Wahrscheinlichkeits 26, 157-170.
- [13] Harmuth, H. (1972). Transmission of Information by Orthogonal Functions, Springer-Verlag, Berlin.
- [14] Kohn, R. (1980 a). On the Spectral Decomposition of Stationary Time Series Using Walsh Functions, I. Adv. Appl. Prob. 12, 183-199.
- [15] Kohn R. (1980 b). On the Spectral Decomposition of Stationary Time Series Using Walsh Functions, II. Adv. Appl. Prob. 12, 462-474.
- [16] Jacobs, P. A. and Lewis, P. A. W. (1978 a). Discrete Time Series Generated by Mixtures I: Correlational and Runs Properties, J. R. Statist. Soc. B 40 (1), 94-105.

- [17] Jacobs, P. A. and Lewis, P. A. W. (1978 b). Discrete Time Series Generated by Mixtures II: Asymptotic Properties, J. R. Statist. Soc. B 40(2), 222-228.
- [18] Jacobs, P. A. and Lewis, P. A. W. (1983). Stationary Discrete Autoregressive Moving Average Time Series Generated by Mixtures, J. Time Series Anal. 4, 18-36.
- [19] Langberg, N. A. and Stoffer, D. S. (1985). Moving Average Models with Bivariate Exponential and Geometric Distributions, Technical Report No. 85-07, Center for Multivariate Analysis, University of Pittsburgh.
- [20] Lewis, P. A. W. (1980). Simple Models for Positive-Valued and Discrete-Valued Time Series with ARMA Correlation Structure, Multivariate Analysis V, ed. by P. R. Krishnaiah, North-Holland, 151-166.
- [21] Morettin, P. A. (1974 a). Limit Theorems for Stationary and Dyadic-Stationary Processes, Biol. Soc. Brasil. Mat. 5, 97-104.
- [22] Morettin, P. A. (1974 b). Walsh-Function Analysis of a Certain Class of Time Series, Stoch. Processes and Their Appl. 2, 183-193.
- [23] Morettin, P. A. (1981). Walsh Spectral Analysis, SIAM Review 23, 279-291.
- [24] Morettin, P. A. (1983). A Note on a Central Limit Theorem for Stationary Processes, J. Time Series Anal. 4, 49-52.
- [25] Ott, J. and Kronmal, R. A. (1976). Some Classification Procedures for Multivariate Binary Data Using Orthogonal Functions, J. Amer. Statist. Assoc. 71, 391-399.
- [26] Robinson, G. S. (1972). Discrete Walsh and Fourier Power Spectra, Proc. Symp. Applications of Walsh Functions, Washington, DC, 298-309.
- [27] Scheffé, H. (1959). The Analysis of Variance, Wiley, New York.
- [28] Skorokhod, A. V. (1956). Limit Theorems for Stochastic Processes, Theory Prob. Appl. 1, 261-290.
- [29] Stoffer, D. S. (1985). Central Limit Theorems for Finite Walsh-Fourier Transforms of Weakly Stationary Time Series, to appear in J. Time Series Anal.
- [30] Stoffer, D. S. and Panchalingam, T. (1985). A Walsh-Fourier Approach to the Analysis of Binary Time Series, to appear in: Proceedings of the Joshi Festschrift, University of Western Ontario, London, Ontario, May 1985.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 86-0064	2. GOVT ACCESSION NO. ADA 166 139	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Walsh-Fourier Analysis of Discrete-Valued Time Series		5. TYPE OF REPORT & PERIOD COVERED Technical - November 1985
7. AUTHOR(s) David S. Stoffer		6. PERFORMING ORG. REPORT NUMBER 85-46
9. PERFORMING ORGANIZATION NAME AND ADDRESS Center for Multivariate Analysis 515 Thackeray Hall University of Pittsburgh, Pittsburgh, PA		8. CONTRACT OR GRANT NUMBER(s) F49620-85-C-0008 AFOSR-84-0113
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research Department of the Air Force Bolling Air Force Base, DC 20332		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS G1102F 2304 / A5
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		12. REPORT DATE November 1985
		13. NUMBER OF PAGES 40
		15. SECURITY CLASS. (of this report) Unclassified
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited		18a. DECLASSIFICATION/DOWNGRADING SCHEDULE
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Discrete-valued time series, Walsh-Fourier spectrum, Walsh-Fourier transform, discrete signal-plus-noise models, regression and analysis of power, Markov chains.		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) An approach to the analyses of discrete-valued time series is discussed. The analyses are accomplished in the spectral domain using the Walsh-Fourier transform which is based on Walsh functions. This approach will enable an investigator of discrete systems to analyze the data in terms of square-waveforms and sequence rather than sine-waves and frequency. We develop a general signal-plus-noise type model for discrete-		

Unclassified

CLASSIFICATION OF THIS PAGE(When Data Entered)

valued time series in which Walsh-Fourier spectral analysis is of interest. We consider the problems of detecting whether or not a common signal exists in repeated measures on discrete-valued time series and in discrete-valued processes collected in an experimental design. We show that these models may depend on unknown regression parameters and we develop consistent estimates of these parameters based on the finite Walsh-Fourier transform. Applications to certain Markov models are given, however, the methods presented also apply to non-Markov cases.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

END

Dtic

5-86